

DÉBATS

« Il faut ramener l'activité sondagère dans le champ scientifique »

TRIBUNE

Michel Lejeune

Statisticien

Avec l'apparition d'Internet, les sociétés de sondage se sont peu à peu éloignées de la rigueur scientifique. Pour qu'elles la retrouvent, les commanditaires doivent imposer un standard de charte sur les informations à livrer, afin que les experts puissent effectuer une analyse de qualité, estime le statisticien Michel Lejeune, dans une tribune au « Monde ».

Publié le 14 décembre 2021 à 19h00 | Lecture 6 min.

Article réservé aux abonnés

Tribune. L'article récent de Luc Bronner (« Dans la fabrique opaque des sondages », [Le Monde](#) du 5 novembre 2021) a dû retourner bien des esprits. Il a donné la preuve évidente que les sondages d'opinion étaient loin d'avoir la qualité vantée par les sondeurs, aussi bien sur le plan de la sélection et du contrôle des échantillons que sur la nature des questionnaires. Pourtant, combien d'observateurs avertis n'ont-ils pas alertés sur les biais de sélection des access panels Internet devenus la norme ? On les a écoutés distraitemment en revenant toujours vers la parole rassurante des sondeurs.

Avec l'apparition et l'usage d'Internet, les sociétés de sondage se sont détournées de la méthode scientifique. Dans leur positionnement concurrentiel, elles ont toutes cédé aux avantages des faibles coûts, de l'immédiat et de la prolifération aux dépens de la qualité. Elles ont développé une démarche empirique, laquelle, n'ayant pas de justification théorique, les a amenées à construire un discours pseudoscientifique séduisant mais peuplé de contre-vérités. Certes cette démarche peut, de temps à autre, donner des résultats jugés corrects dans les intentions de vote proches de la date du scrutin, mais elle n'offre aucune garantie a priori.

Un concept flou

Trois types de contre-vérités constituent le socle de la communication des sondeurs. Ils concernent la représentativité de leurs échantillons, les redressements et l'analyse des résultats. Examinons-les successivement.

Tous les sondages sont déclarés « représentatifs par la méthode des quotas ». Premier problème, au-delà de son sens historique très spécifique, il n'existe pas de définition de la représentativité. Il serait intéressant de poser la question de sa signification à ceux qui se réfèrent à ce terme. Pour chacun d'entre nous, c'est un concept flou qui assurerait que l'on puisse faire de bonnes estimations sur la population.

La formulation des sondeurs fait accroire que cela est garanti par la « méthode des quotas » du fait qu'elle produit, selon leurs termes, un « modèle réduit de la population ». Or ce modèle ne s'applique qu'à quatre critères sociodémographiques de quotas (âge, genre, profession, zone de résidence), ce qui est totalement insuffisant pour fournir de bonnes estimations.

Ainsi, comme l'a montré concrètement l'enquête du *Monde*, les échantillons issus des « access panels » [*bases de données*] n'ont pas cette qualité, quand bien même ils sont construits sur des quotas. D'ailleurs, le seul fait que les panélistes doivent être des internautes coutumiers, excluant d'emblée de l'ordre de 40 % de la population, selon les études de l'Insee, suffirait à le prouver.

Contre-vérité flagrante

La vérité scientifique est que seule une sélection aléatoire, complétée ou non par un calage sur des quotas, permet d'opérer des estimations sans biais, quelles qu'elles soient. Cela pourrait être une définition de la représentativité. Or les sondeurs rechignent à reconnaître les vertus de l'aléatoire. Déclarer qu'un échantillon est représentatif du simple fait qu'il répond à des quotas est une contre-vérité flagrante.

Alors qu'ils sont parvenus à convaincre le grand public qu'ils utilisent des méthodes scientifiques éprouvées, ils n'hésitent pas à déclarer que les résultats « bruts » de leurs sondages sont mauvais, voire « faux », selon [*l'ancien sondeur*] Roland Cayrol. Ils avancent alors que la technique du redressement va permettre d'y remédier. Celle-ci occupe effectivement une place dans la panoplie scientifique mais en suivant des principes précis. Le problème, ici, est que les sondeurs utilisent les redressements de façon empirique, sans réel discernement.

Archives : [Sondages : « Des biais affectent la qualité et la représentativité des échantillons »](#)

Dans les sondages d'intention de vote, on recueille les déclarations des répondants à propos de leurs votes lors d'élections antérieures dans le but de rectifier l'échantillon en le calant sur les résultats réels de ces élections. Ce qui peut paraître sensé. Ce type de redressement a été adoubi de façon circonstancielle pour traiter le problème de la très forte sous-estimation du vote en faveur de Jean-Marie Le Pen.

Mécaniquement, il est vrai que les redressements sur les résultats d'élections antérieures remontaient son score, mais dans une ampleur très hasardeuse en raison de l'étonnant manque de fiabilité des déclarations des répondants, y compris sur des élections proches.

Commentaire abusif

Aujourd'hui, le vent a tourné et le vote RN est plutôt surestimé par les « access panels ». J'ai montré dans un article sur [AOC](#) que, lors des dernières élections régionales, ces redressements sont venus encore aggraver cette surestimation pour atteindre des taux record. Dire que les redressements utilisés viennent corriger les biais, une conception bien ancrée dans le public, est une deuxième contre-vérité.

L'analyse des résultats est également truffée de fausses vérités par le fait qu'elle commente abusivement des écarts d'un ou deux points qui n'ont aucune signification statistique en vertu des marges d'erreur. Les prestataires de sondages, certes, produisent un tableau global résumant ces marges, mais n'en tiennent jamais compte.

Archives : [« Pouvoir se fier à des sondages sérieux reste un enjeu démocratique »](#)

Ainsi au moment où la pression sur cette exigence se fait plus forte, on voit poindre une contre-vérité connexe dans l'argumentaire défensif, à savoir que les sondages sur quotas ne se prêteraient pas aux

marges d'erreur. Cela revient donc à dire que l'on n'a aucune garantie de précision et devrait les disqualifier comme instruments de mesure. En réalité cela est faux car les marges d'erreur sont une référence pour un échantillonnage idéal et peuvent servir d'approximation optimiste pour les échantillons réels.

La liste des présentations trompeuses ne s'arrête pas là. J'aimerais encore mentionner la rhétorique d'un contrôle très strict de la part de la Commission des sondages alors que, dans la réalité, la plupart des directives ont pu être habilement contournées.

Un lobby

Bien des dérives ont pu perdurer grâce à l'opacité de la production des chiffres publiés que l'on justifie par un secret professionnel. Ce n'est qu'un mythe car les sociétés sont un lobby au sein duquel s'échangent les personnes et les informations. Les pratiques sont identiques partout, ne se distinguant que sur quelques variantes bien connues.

En fait, ce supposé secret vise les destinataires des études ; ce qui n'a pas lieu d'être dans un contexte prétendument scientifique, comme le martèle le sénateur socialiste du Loiret Jean-Pierre Sueur dans son combat pour la transparence. En disposant de très peu d'informations, il est très difficile de débusquer les failles. Personnellement, les conclusions parmi les plus cruciales que j'ai pu tirer l'ont été grâce à la publication des données brutes et des variantes de redressement par la seule société ouverte à la transparence (Opinionway).

Les commanditaires de sondages ont le pouvoir dès aujourd'hui d'imposer à leurs prestataires un standard de charte sur les informations à livrer avec chaque sondage afin que les différents observateurs avisés, experts internes comme externes, puissent effectuer proprement une analyse de qualité.

Voici, pour exemple, quelques exigences phares pouvant être posées immédiatement : identification du ou des panels utilisés, mesures de contrôle et gestion des panélistes, mesures de contrôle des informations recueillies auprès des répondants, profils sociologiques ou autres des panélistes en général et des répondants au sondage en particulier, questionnaire complet expédié, publication des chiffres bruts, publication des résultats intégrant systématiquement les marges d'erreur sous une forme analogue au graphe ci-joint.

Si l'on veut stopper le déploiement médiatique anarchique et redonner la confiance dans les sondages, il faudra ramener l'activité sondagière dans le champ scientifique en étant honnête sur la nature des sondages et dans l'acceptation de leurs limites. Face aux « access panels » actuels, il existe des solutions pratiquées par des instituts non commerciaux mais avec des budgets temps et financiers importants. Cependant, l'élévation des coûts serait aussi une mesure pour le retour à la raison, notamment en limitant le nombre de sondages et le nombre de prestataires capables de se remettre en question.

¶ **Michel Lejeune** est professeur honoraire de l'université Grenoble-Alpes. Il a été durant vingt ans membre du comité scientifique du Centre d'étude des supports de publicité (CESP). Il est auteur de « La Singulière Fabrique des sondages d'opinion » (L'Harmattan, 168 pages, 19 euros).

Michel Lejeune (Statisticien)

Services